



## Appraisal

## Research Note: Comparing interventions with network meta-analysis

Physiotherapists have become familiar with using research findings to inform practice. A physiotherapist who is interested in using research findings to determine the best physiotherapy intervention for a patient with osteoarthritis of the knee might look for high-quality systematic reviews of randomised, controlled trials (RCTs). A search of PEDro (<https://www.pedro.org.au>) would identify a Cochrane systematic review<sup>1</sup> on the efficacy of exercise for patients with osteoarthritis of the knee. The review provides high-quality evidence that exercise reduces pain and moderate-quality evidence that exercise improves physical function when compared with not exercising.

However, the results of a systematic review that compares only two interventions may not provide all the information required to make a clinical decision. The clinician, for example, may be interested in discriminating between flexibility and aerobic exercise for osteoarthritis. The osteoarthritis systematic review is not helpful for this purpose because it excluded RCTs that compared exercise interventions with each other. To get an answer, the clinician may compare the benefit of flexibility exercise compared with no exercise to the benefit of aerobic exercise compared with no exercise. Such an indirect comparison, which is not obtained from a single RCT, is valid under certain assumptions.<sup>2–5</sup> If these assumptions are not fulfilled, the indirect comparison may lead to biased conclusions.

Based on the notion of indirect comparison, a new type of evidence synthesis has emerged over the last 20 years to address this challenge. Network meta-analysis compares several available interventions for a clinical problem with each other in a single meta-analysis. The analysed comparisons may include comparisons between interventions that have not been directly compared in RCTs.<sup>2,5–9</sup> A league table of the interventions in the network, displaying all possible pairwise relative effect sizes, can be used to compare and rank the interventions according to efficacy, acceptability or safety. This provides critical information to inform clinical decision-making.

The field of network meta-analysis has developed rapidly.<sup>10–13</sup> Network meta-analysis has far-reaching relevance to clinicians, researchers, guideline-developers, regulators and policy-makers.<sup>14–16</sup> The purpose of this Research Note is to provide a user-friendly overview of the principles and assumptions that underlie network meta-analysis.

### What is a network of interventions?

The central element of network meta-analysis is the network, which distinguishes it from conventional pairwise meta-analyses. Networks display the evidence of the effectiveness of interventions for a clinical condition. For example, a network might display evidence of the effectiveness of exercise interventions for hip or knee osteoarthritis,<sup>17</sup> non-pharmacological interventions for cancer-related fatigue,<sup>18</sup> or non-operative treatment for chronic calcific shoulder tendinopathy.<sup>19</sup> A network consists of nodes and edges. Each node in the network represents an intervention. Each

edge (a line between two nodes) represents a comparison between two interventions that has been evaluated in at least one RCT. If there is no edge between two nodes, no data from RCTs have compared these interventions. In this way, the network displays all the available comparisons in the evidence base and alludes to all the possible comparisons.

### Network construction

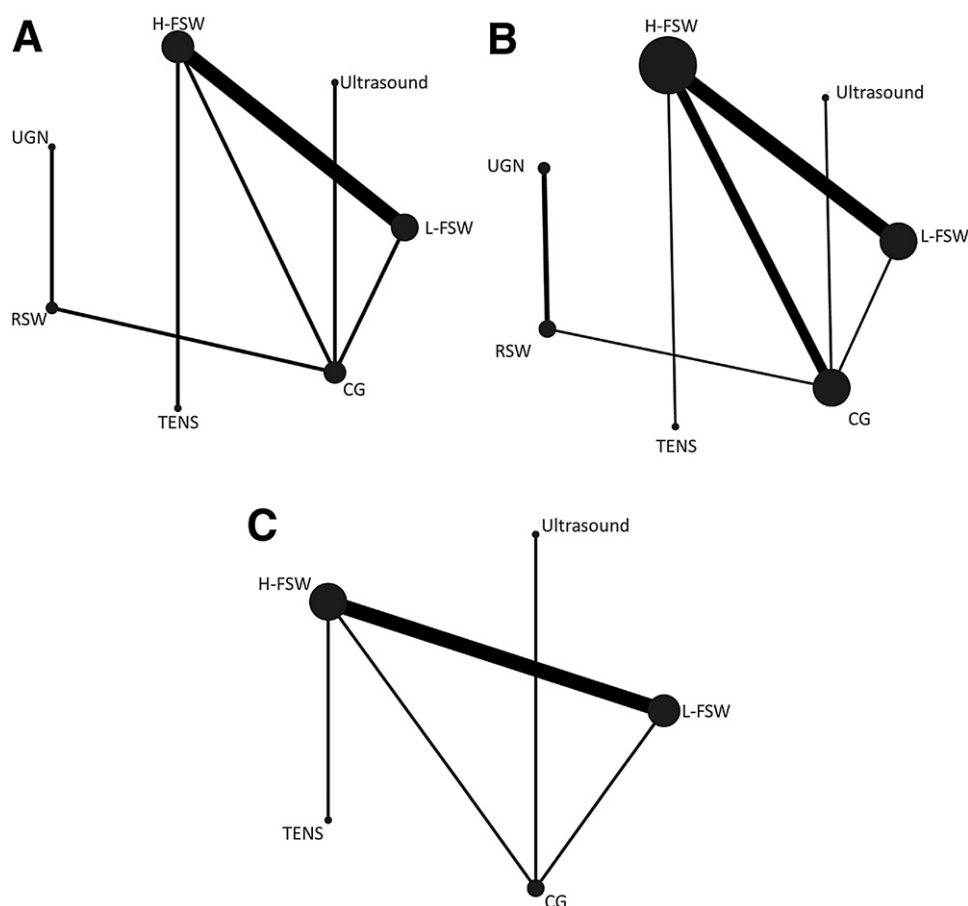
Multiple factors contribute to network construction. For researchers conducting a network meta-analysis, a key step is determining the PICOS (Participants, Interventions, Comparisons, Outcomes and Study types). The descriptions of the participants (P) and the comparisons (C) to be studied influence which RCTs will be included and, hence, the presence and size of the edges in the network. The interventions (I) are the nodes in the network. The selection of outcomes (O) is important, as trials without certain outcomes will not be included in the network meta-analysis for that outcome. Network meta-analysis is usually performed using RCTs (S), although methods exist to incorporate non-randomised data.<sup>20</sup> In the networks of interventions for chronic calcific shoulder tendinopathy,<sup>19</sup> the width of each edge reflects the number of trials for that comparison, each node represents an intervention, and the node diameter is proportional to the number of participants allocated to that treatment. It is also shown that the network structure changes according to outcome (Figure 1).

### The multiple treatment comparison

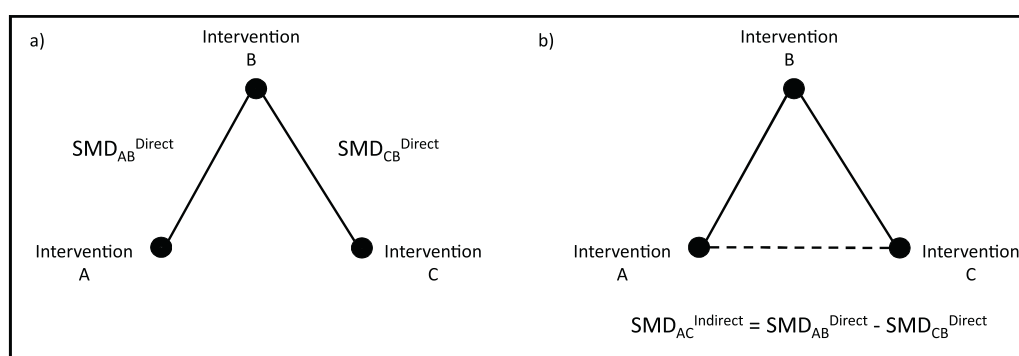
The network displays the number of available *direct* comparisons; comparisons for which there are RCT data available. The absent comparisons, termed *indirect* comparisons, have no RCT data available. The relative treatment effects for these comparisons will be indirectly estimated in network meta-analysis. A systematic review is critical to ensure that all the available direct comparisons are included in the network. Clinicians reading a network meta-analysis article should be satisfied that a rigorous systematic review was performed. There are several guides available for this purpose.<sup>21–24</sup>

### Making a comparison between interventions in a network

The idea of combining *indirect* with *direct* evidence (when the latter is available) characterises network meta-analysis. To illustrate this, a hypothetical example is used. Three interventions form the simplest possible network of three nodes (Figure 2a). A clinician may want to know the effects of these three interventions compared with each other on a continuous outcome, so that the most effective intervention can be provided to a patient. Studies comparing A with B (AB), when synthesised, would produce a standardised mean difference for the direct comparison AB ( $SMD_{AB}^{Direct}$ ) and studies comparing C with B (CB) would produce  $SMD_{CB}^{Direct}$ . If the comparison between A and C



**Figure 1.** Network plots of interventions for chronic calcific shoulder tendinopathy, displayed according to three outcomes (reproduction from Wu et al.<sup>19</sup>). Reproduced with permission. Figure label size increased from original. Original material available at doi: [10.1016/j.apmr.2017.02.030](https://doi.org/10.1016/j.apmr.2017.02.030). Copyright 2017 by the American Congress of Rehabilitation Medicine. UGN = ultrasound-guided needling, RSW = radial extracorporeal shockwave therapy, H-FSW = high-energy focused extracorporeal shockwave therapy, TENS = transcutaneous electrical nerve stimulation, CG = control group (sham treatment or physiotherapy alone), L-FSW = low-energy focused extracorporeal shockwave therapy.



**Figure 2.** A simple three-node network comparing three interventions. SMD = standardised mean difference.

(AC) has not been tested in a clinical trial, as Figure 2a indicates, the network meta-analysis model estimates the ‘missing’ relative treatment effect of AC by using the AB and CB data to estimate an indirect effect of AC as  $SMD_{AC}^{Indirect} = SMD_{AB}^{Direct} - SMD_{CB}^{Direct}$  (Figure 2b). This is called an indirect treatment comparison.

If there are studies directly comparing intervention A and C, their synthesis will provide a  $SMD_{AC}^{Direct}$ . The two estimates,  $SMD_{AC}^{Direct}$  and  $SMD_{AC}^{Indirect}$  can be synthesised as a weighted average to provide  $SMD_{AC}^{Mixed}$ . This is called a mixed treatment comparison. In more complex network structures, all of the other direct comparisons in the network will contribute information to this estimate.<sup>25–27</sup> Thus, network meta-analysis estimates are weighted sums of all direct and indirect comparisons present in

the network. The weighting is influenced by precision, as in pair-wise meta-analysis, and network structure.<sup>25,27</sup>

#### Assumptions underlying indirect comparison and network meta-analysis

All statistical models require assumptions about the data and the underlying parameters. The validity of network meta-analysis depends on the assumption of transitivity. A joint synthesis of the data in the network is valid only if the included studies are similar in all important characteristics except for the interventions being tested.<sup>5,11</sup> This is also equivalent to the assumption that a participant included in any trial could, in principle, be randomised

to receive any of the interventions in a network. In the hypothetical example (Figure 2), a valid estimation of  $SMD_{AC}^{Mixed}$  requires that the three groups of studies – AB, CB and AC – are similar with respect to relevant effect modifiers (characteristics that might alter the relative treatment effects).

Transitivity is not a binary construct – present or not – there will be a degree of confidence in whether the assumption has been met. Confidence in the assumption is important because it reflects the extent to which the results from the evidence in the network are likely to be free from confounding (eg, see references<sup>28–30</sup>). Researchers conducting a network meta-analysis should think carefully about the process they will use to ensure that the transitivity assumption is likely to hold. This should begin at protocol stage<sup>11</sup> where, again, setting the PICOS is important. Setting narrow criteria for the eligible participants, study settings and included interventions make the transitivity assumption more plausible, yet may result in networks that are sparse or disconnected and cannot be analysed. In the protocol, potential effect modifiers should be listed. Then, in the review, the distribution of the effect modifiers across the network should be assessed and described. The interested reader is referred to descriptions of how to assess transitivity from the perspectives of a user of network meta-analysis<sup>21,24</sup> and a researcher conducting network meta-analysis.<sup>5,10,11,26</sup>

The network meta-analyses for chronic calcific shoulder tendinopathy<sup>19</sup> excluded rotator cuff tears. This is an example of a relatively narrow definition of inclusion criteria. It is possible that people with rotator cuff tears respond differently to the interventions studied in the network meta-analysis. The presence of the covariate (rotator cuff tear) may confound the network meta-analysis estimates if the covariate is not equally distributed across all the comparisons in the network. The narrow inclusion criteria prevent this altogether by excluding the covariate from the network.

Hilfiker et al assessed the distribution of the covariate ‘allocation concealment’ across the networks of interventions for cancer-related fatigue.<sup>18</sup> Inadequate allocation concealment is associated with trial effect estimates.<sup>31,32</sup> In a network meta-analysis, this will introduce bias if the covariate is not evenly distributed across the studies

grouped by comparison. A balanced distribution controls for the influence of the covariate. Methods for assessing the distribution of covariates are described in two studies.<sup>26,33</sup>

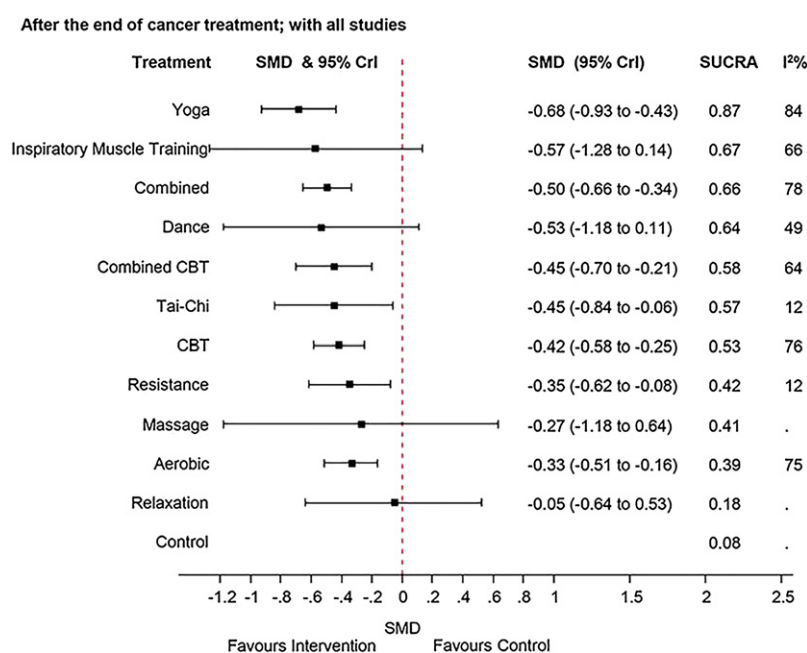
### Heterogeneity and inconsistency in network meta-analysis

As in conventional meta-analysis, heterogeneity is an important consideration. However, there is a further consideration in network meta-analysis: inconsistency. Inconsistency and heterogeneity are inter-related.<sup>5</sup> Both should be considered and the results should be interpreted after accounting for their magnitudes. Here, the next section focuses on describing the simplest form of inconsistency and the reader is referred for further detail to the technical literature.<sup>10,11,34,35</sup>

Inconsistency is associated with violation of the transitivity assumption and could be described as a signal in the data that the various pieces of evidence – direct and indirect – do not fit together. In its simplest form, inconsistency is present when the direct estimate for a comparison (eg,  $SMD_{AC}^{Direct}$ ) differs beyond chance from the indirect estimate (eg,  $SMD_{AC}^{Indirect}$ ). It follows that a statistical evaluation of inconsistency is possible only in the presence of direct and indirect evidence for the same comparison.<sup>10</sup> An inconsistency factor can be estimated as  $SMD_{AC}^{Direct} - SMD_{AC}^{Indirect}$  and the assumption that inconsistency is 0 can be tested using a Z-test.<sup>4,36</sup>

It is important to know that the magnitude of heterogeneity – the method used to estimate it in the network meta-analysis model, researcher analytic choices<sup>37–39</sup> and covariate values in advanced models<sup>40</sup> – will influence detection of inconsistency. The low power of the tests for inconsistency in many applications<sup>27,37,38,41</sup> means that a non-significant test does not guarantee that the assumptions underlying network meta-analysis are met. Clinicians reading network meta-analysis articles should make sure that several tests and measures of inconsistency<sup>10,11,42</sup> are performed and described. Researchers should follow these practices<sup>43</sup> and interpret results with caution, particularly if the number of studies is small.

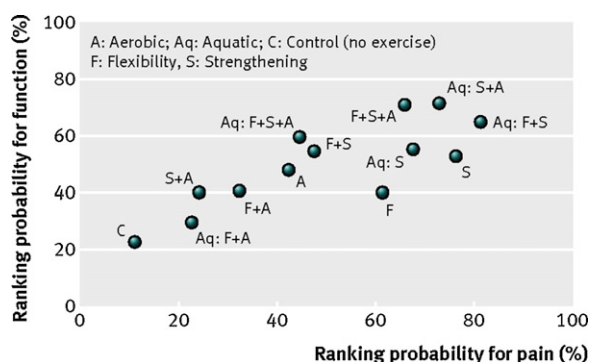
Of the current examples, the network meta-analysis of interventions for chronic calcific shoulder tendinopathy<sup>19</sup> used



**Figure 3.** Network meta-analysis summary standardised mean differences (SMD); non-pharmaceutical treatments versus control (usual care) for cancer-related fatigue (reproduction from Hilfiker et al<sup>18</sup>).

Original material available under Public License (CC BY-NC 4.0) at doi: [10.1136/bjsports-2016-096422](https://doi.org/10.1136/bjsports-2016-096422). Copyright article author (or their employer) 2017. Material has been adapted (figure cropped and title changed) from the original.

CBT = cognitive behavioural therapy, CrI = credible interval, SMD = standardised mean difference, SUCRA = surface under the cumulative ranking curve.



**Figure 4.** Surface under the cumulative ranking curve (SUCRA % values) for pain and function for exercise treatments for knee or hip osteoarthritis (reproduction Uthman et al.<sup>17</sup>).

Original material available under Public License (CC BY-NC 3.0) at doi: [10.1136/bmj.f5555](https://doi.org/10.1136/bmj.f5555). Title has been changed from the original.

the recommended approach of applying several tests of inconsistency. The authors used an omnibus test of inconsistency (the design-by-treatment test<sup>41</sup>), as well as the 'loop-specific'<sup>4</sup> and 'node-splitting'<sup>44</sup> approaches to assess inconsistency associated with specific comparisons. Hilfiker et al used the node-splitting model in their analysis of interventions for cancer-related fatigue,<sup>18</sup> and Uthman et al used the loop-specific approach in their analysis of exercise interventions for knee or hip osteoarthritis.<sup>17</sup> Single tests and measures are not ideal because the high heterogeneity observed in these networks may cause spurious results.

### Why perform a network meta-analysis?

Network meta-analysis is a statistical method for comparing multiple treatments. The formalised process of making an indirect comparison between interventions – particularly the specification and evaluation of assumptions – conveys the extent to which the results from the joint synthesis of evidence could be subject to bias. Network meta-analysis uses the entire evidence base to construct the estimates of effect, which may reduce research waste. Each network meta-analysis estimate incorporates direct and indirect evidence and, hence, is more precise than estimates from direct studies alone.<sup>3,7,8,45–49</sup>

Synthesis of the entire evidence base with network meta-analysis is advantageous for decision-making. The availability of effect estimates for each comparison is informative; however, an added advantage of network meta-analysis is the establishment of a hierarchy of interventions. Figure 3 displays a network meta-analysis forest plot, which in contrast to a conventional forest plot displays the effect size for each intervention compared to a reference. The effect sizes for all comparisons are listed in league tables, for example.<sup>17</sup> Ranking probabilities can be used to produce probabilistic statements about the relative advantage of an intervention compared to all other alternatives (the SUCRA in Figure 3). Often, these are presented for multiple relevant outcomes, for example, the effect of exercise on pain and on function (Figure 4).

### Limitations of network meta-analysis?

Network meta-analysis can now be fitted in standard statistical software, such as Stata and R and is becoming increasingly popular. For advanced, related, techniques like network meta-regression, which are often needed in applications, it is necessary to employ more flexible models and techniques that are not accessible to most non-statisticians (eg, hierarchical models fitted in a Bayesian framework). Nevertheless, the main challenge in the process is the evaluation of the assumption underlying network meta-analysis.

Transitivity is a significant and often strong assumption. Whilst a robust argument can often be made for transitivity, it can never be proven because participants cannot actually be randomised within the indirect comparisons. This means that despite rigorous attention, for example using directed acyclic graphs,<sup>50</sup> unobserved confounding may still bias the estimates.<sup>5,50</sup>

### Summary

Network meta-analysis is the 'next generation evidence synthesis tool'.<sup>11</sup> Physiotherapy practice and research stands to benefit from the powerful features of the methodology. Yet, network meta-analysis relies on key assumptions in study design and analysis. Physiotherapists should understand these assumptions and ensure they are met in the network meta-analyses that they use or perform.

**Competing interests:** MKB is supported by an Australian Government Research Training Program Scholarship, a UNSW Research Excellence Award and a NeuRA PhD Candidature Supplementary Scholarship. GS is a Marie-Skłodowska Curie fellow (MSCA-IF-703254).

**Sources of support:** Nil.

**Acknowledgements:** Nil.

**Correspondence:** Matthew Bagg, Pain Research, Education and Management Program, Neuroscience Research Australia, Sydney, Australia. Email: [m.bagg@neura.edu.au](mailto:m.bagg@neura.edu.au)

**Matthew K Bagg<sup>a,b</sup>, Georgia Salanti<sup>c</sup> and James H McAuley<sup>a,d</sup>**

<sup>a</sup>Neuroscience Research Australia

<sup>b</sup>Prince of Wales Clinical School & New College Village, University of New South Wales, Sydney, Australia

<sup>c</sup>Institute for Social and Preventive Medicine, University of Bern, Bern, Switzerland

<sup>d</sup>School of Medical Sciences, University of New South Wales, Sydney, Australia

### References

- Fransen M, et al. *Cochrane Database Syst Rev.* 2015;1:CD004376.
- Song F, et al. *BMJ.* 2003;326(7387):472.
- Song F, et al. *J Clin Epidemiol.* 2008;61:455–463.
- Bucher HC, et al. *J Clin Epidemiol.* 1997;50:683–691.
- Jansen JP, Naci H. *BMC Med.* 2013;11:159.
- Caldwell DM, et al. *BMJ.* 2005;331(7521):897–900.
- Jansen JP, et al. *Value Heal.* 2011;14:417–428.
- Mills EJ, et al. *BMJ.* 2013;2914:10–15.
- Sutton A, et al. *Pharmacoeconomics.* 2008;26:753–767.
- Efthimiou O, et al. *Res Synth Methods.* 2016;7:236–263.
- Salanti G. *Res Synth Methods.* 2012;3:80–97.
- Petropoulou M, et al. *J Clin Epidemiol.* 2017;82:20–28.
- Zarin W, et al. *BMC Med.* 2017;15:3.
- Naci H, O'Connor AB. *J Clin Epidemiol.* 2013;66:812–816.
- Leucht S, et al. *Eur Arch Psychiatry Clin Neurosci.* 2016;266:477–480.
- Rouse B, et al. *Ann Intern Med.* 2016;164:674–682.
- Uthman OA, et al. *BMJ.* 2013;347:f5555.
- Hilfiker R, et al. *Br J Sports Med.* 2017. (online first, 13 May 2017).
- Wu Y, et al. *Arch Phys Med Rehabil.* 2017;98:1678–1692.
- Efthimiou O, et al. *Stat Med.* 2017;36:1210–1226.
- Jansen JP, et al. *Value Heal.* 2014;17:157–173.
- Ades AE, et al. *Med Decis Mak.* 2013;33:679–691.
- Chaudhry H, et al. *Clin Orthop Relat Res.* 2015;473:2172–2178.
- Footo CJ, et al. *Clin Orthop Relat Res.* 2015;473:2166–2171.
- Lu G, et al. *Res Synth Methods.* 2011;2:43–60.
- Chaimani A, et al. *PLoS One.* 2013;8:e76654.
- Krahn U, et al. *BMC Med Res Methodol.* 2013;13:35.
- Guo J-R, et al. *Ann Pharmacother.* 2017;51:1041–1052.
- Guo J-R, et al. *Ann Pharmacother.* 2017;52:99.
- Bagg MK, McAuley JH. *Ann Pharmacother.* 2018;52:97–98.
- Dechartres A, et al. *J Clin Epidemiol.* 2016;77:24–37.
- Savovic J, et al. *Ann Intern Med.* 2012;157:429–438.
- Batson S, et al. *J Clin Epidemiol.* 2017;86:182–195.
- Dias S, et al. *Med Decis Mak.* 2013;33:641–656.
- Dias S, et al. *Med Decis Mak.* 2013;33:618–640.
- Lu G, Ades AE. *J Am Stat Assoc.* 2006;101:474.
- Song F, et al. *BMC Med Res Methodol.* 2012;12:138.
- Veroniki AA, et al. *BMC Med Res Methodol.* 2014;106:12.
- Veroniki AA, et al. *Int J Epidemiol.* 2013;42:332–345.
- Donegan S, et al. *Res Synth Methods.* 2017;8:485–495.

41. Higgins JPT, et al. *Res Synth Methods*. 2012;3:98–110.
42. Cipriani A, et al. *Ann Intern Med*. 2013;159:130–137.
43. Mills EJ, et al. *JAMA*. 2012;308:1246–1253.
44. Dias S, et al. *Stat Med*. 2010;29:932–944.
45. Lu G, Ades AE. *Stat Med*. 2004;23:3105–3124.
46. Madan J, et al. *Value Heal*. 2011;14:953–960.
47. Higgins JPT, Whitehead A. *Stat Med*. 1996;15:2733–2749.
48. Salanti G, et al. *Stat Methods Med Res*. 2008;17:279–301.
49. Lumley T. *Stat Med*. 2002;21:2313–2324.
50. Jansen JP, et al. *J Clin Epidemiol*. 2012;65:798–807.